

AMENDMENT TO THE CLAIMS

1. (currently amended) A computer readable storage media storing instructions readable by a computer which, when implemented, cause the computer to perform a method comprising:

segmenting a sentence of Chinese characters into constituent Chinese words having one or more Chinese characters;

recognizing an overlapping ambiguity string in the segmented sentence, wherein the overlapping ambiguity string comprises at least three Chinese characters having at least two possible segmentations, wherein each possible segmentation comprises a right portion and a left portion;

obtaining probability information for each possible segmentation, wherein the probability information is based on at least one context feature adjacent the overlapping ambiguity string and one of the right portion or left portion of the possible segmentation, and wherein the at least one context feature comprises a Chinese character; and

outputting an indication for selecting one of the at least two possible segmentations as a function of the obtained probability information.

2. (currently amended) The computer readable storage media of claim 1, wherein obtaining the probability information comprises obtaining probability information from a language model ~~based on the at least one context feature and a left or right portion of the overlapping ambiguity string.~~

3. (Previously presented) The computer readable storage media of claim 2 wherein the language model comprises a trigram model.

4. (previously presented) The computer readable storage media of claim 2 wherein outputting an indication for selecting one of the at least two possible segmentations comprises classifying the probability information.
5. (Previously presented) The computer readable storage media of claim 4 wherein classifying comprises classifying using Naïve Bayesian Classification.
6. (Previously presented) The computer readable storage media of claim 1 wherein segmenting the sentence comprises performing a Forward Maximum Matching (FMM) segmentation of the input sentence and a Backward Maximum Matching (BMM) segmentation of the input sentence.
7. (Currently amended) The computer readable storage media of claim 6 wherein recognizing the overlapping ambiguity string comprises recognizing a possible segmentation O_f of the overlapping ambiguity string from the FMM segmentation and a possible segmentation O_b of the overlapping ambiguity string from the BMM segmentation.
8. (currently amended) The computer readable storage media of claim 7, wherein outputting the indication comprises selecting one of the at least two possible segmentations as a function of a set of context features surrounding the overlapping ambiguity string.
9. (previously presented) The computer readable storage media of claim 8 wherein the set of context features comprises words or grammatical features surrounding the overlapping ambiguity

string.

10. (previously presented) The computer readable storage media of claim 8, wherein outputting the indication comprises classifying the probability information of the set of context features and O_f .

11. (previously presented) The computer readable storage media of claim 8, wherein outputting the indication comprises classifying the probability information of the set of context features and O_b .

12. (previously presented) The computer readable storage media of claim 8, outputting the indication comprises determining which of O_f or O_b has a higher probability as a function of the set of context features.

13. (cancelled)

14. (currently amended) A method of segmentation of a sentence of Chinese text, the sentence having an overlapping ambiguity string, the method comprising:

generating a Forward Maximum Matching (FMM)
segmentation of the sentence;

generating a Backward Maximum Matching (BMM)
segmentation of the sentence;

recognizing the overlapping ambiguity string based on a
difference between the FMM segmentation and the BMM
segmentation;

obtaining probability information based on at least one
context feature surrounding the overlapping
ambiguity string and at least part of the

overlapping ambiguity string, wherein the at least
one context feature comprises a Chinese character;
and

outputting an indication for selecting one of the FMM
segmentation and the BMM segmentation as a function
of obtained probability information.

15. (previously presented) The method of claim 14 wherein
outputting includes selecting one of the FMM segmentation of the
overlapping ambiguity string and the BMM segmentation of the
overlapping ambiguity string based on higher probability.

16. (previously presented) The method of claim 15 wherein
obtaining probability information comprises using an N-gram
model.

17. (previously presented) The method of claim 16 wherein
obtaining probability information comprises obtaining probability
information about a first word of the overlapping ambiguity
string.

18. (previously presented) The method of claim 16, wherein
obtaining probability information comprises using probability
information about a last word of the overlapping ambiguity
string.

19. (previously presented) The method of claim 16, wherein
obtaining probability information comprises using the N-gram
model that includes probability information for context words
surrounding the overlapping ambiguity string.

20. (previously presented) The method of claim 16, wherein
using the N-gram model comprises using trigram probability

information about a string of words comprising a first word of the overlapping ambiguity string and two context words to the left of the first word.

21. (previously presented) The method of claim 16, wherein using the N-gram model comprises using trigram probability information about a string of words comprising a last word of the overlapping ambiguity string and two context words to the right of the last word.

22. (previously presented) The method of claim 14, wherein outputting includes using Naïve Bayesian Classifiers.

23. (previously presented) The method of claim 14, wherein obtaining probability information comprises obtaining trigram probability information and constructing an ensemble of Naive Bayesian Classifiers from the trigram probability information.

24. (previously presented) The method of claim 23, wherein outputting an indication comprises identifying one of the FMM segmentation and the BMM segmentation based on probability calculated from the ensemble of Naïve Bayesian Classifiers.

25. (currently amended) A method of segmenting a sentence of Chinese text comprising:

recognizing an overlapping ambiguity string in the sentence;

receiving probability information from an N-gram language model comprising probability information for constituent words of the overlapping ambiguity string and context features surrounding the overlapping ambiguity string, wherein the context features comprise at least one Chinese character;

resolving the overlapping ambiguity string based on the received probability information.

26. (previously presented) The method of claim 25, wherein receiving probability information comprises receiving probability information from a trigram language model.

27. (previously presented) The method of claim 25, and further comprising generating an ensemble of classifiers with the received probability information.

28. (previously presented) The method of claim 25, wherein recognizing the overlapping ambiguity string comprises:

generating a Forward Maximum Matching (FMM) segmentation for the sentence;
generating a Backward Maximum Matching (BMM) segmentation for the sentence;
recognizing the overlapping ambiguity string based on the FMM segmentation and the BMM segmentation of the sentence.

29. (original) The method of claim 28 and further comprising generating an ensemble of classifiers as a function of the N-gram model.

30. (previously presented) The method of claim 29 wherein generating the ensemble of classifiers includes approximating probabilities of the FMM and BMM segmentations of the overlapping ambiguity string as being equal to the product of individual unigram probabilities of individual words in the FMM and BMM segmentations of the overlapping ambiguity string.

31. (previously presented) The method of claim 29, wherein generating the ensemble of classifiers includes approximating a joint probability of a set of context features conditioned on an existence of one of the segmentations of the overlapping ambiguity string based on a corresponding probability of a leftmost and a rightmost word of the corresponding overlapping ambiguity string.